SECTION 7

DILUTION WATER

7.1 TYPES OF DILUTION WATER

7.1.1 The type of dilution water used in effluent toxicity tests will depend largely on the objectives of the study.

7.1.1.1 If the objective of the test is to estimate the absolute chronic toxicity of the effluent, a synthetic (standard) dilution water is used. If the test organisms have been cultured in water which is different from the test dilution water, a second set of controls, using culture water, should be included in the test.

7.1.1.2 If the objective of the test is to estimate the chronic toxicity of the effluent in uncontaminated receiving water, the test may be conducted using dilution water consisting of a single grab sample of receiving water (if non-toxic), collected either upstream and outside the influence of the outfall, or with other uncontaminated natural water (ground or surface water) or standard dilution water having approximately the same characteristics (hardness, alkalinity, and conductivity) as the receiving water. Seasonal variations in the quality of receiving waters may affect effluent toxicity. Therefore, the pH, alkalinity, hardness, and conductivity of receiving water samples should be determined before each use. If the test organisms have been cultured in water which is different from the test dilution water, a second set of controls, using culture water, should be included in the test.

7.1.1.3 If the objective of the test is to determine the additive or mitigating effects of the discharge on already contaminated receiving water, the test is performed using dilution water consisting of receiving water collected immediately upstream or outside the influence of the outfall. A second set of controls, using culture water, should be included in the test.

7.1.2 An acceptable dilution water is one which is appropriate for the objectives of the test; supports adequate performance of the test organisms with respect to survival, growth, reproduction, or other responses that may be measured in the test (i.e., consistently meets test acceptability criteria for control responses); is consistent in quality; and does not contain contaminants that could produce toxicity. Receiving waters, synthetic waters, or synthetic waters adjusted to approximate receiving water characteristics may be used for dilution provided that the water meets the above listed qualifications for an acceptable dilution water. USEPA (2000a) provides additional guidance on selecting appropriate dilution waters.

7.1.3 When dual controls (one control using culture water and one control using dilution water) are used (see Subsections 7.1.1.1 - 7.1.1.3 above), the dilution water control should be used to determine test acceptability. It is also the dilution water control that should be compared to effluent treatments in the calculation and reporting of test results. The culture water control should be used to evaluate the appropriateness of the dilution water source. Significant differences between organism responses in culture water and dilution water controls could indicate toxicity in the dilution water and may suggest an alternative dilution water source. USEPA (2000a) provides additional guidance on dual controls.

7.2 STANDARD, SYNTHETIC DILUTION WATER

7.2.1 Standard, synthetic dilution water is prepared with deionized water and reagent grade chemicals or mineral water (Tables 3 and 4). The source water for the deionizer can be ground water or tap water.

7.2.2 DEIONIZED WATER USED TO PREPARE STANDARD, SYNTHETIC, DILUTION WATER

7.2.2.1 Deionized water is obtained from a MILLIPORE® MILLI-Q®, MILLIPORE® QPAK™$_2$ or equivalent system. It is advisable to provide a preconditioned (deionized) feed water by using a Culligan®, Continental®, or

equivalent system in front of the MILLIPORE® System to extend the life of the MILLIPORE® cartridges (see Section 5, Facilities, Equipment, and Supplies).

7.2.2.2 The recommended order of the cartridges in a four-cartridge deionizer (i.e., MILLI-Q® System or equivalent) is (1) ion exchange, (2) ion exchange, (3) carbon, and (4) organic cleanup (such as ORGANEX-Q®, or equivalent) followed by a final bacteria filter. The QPAK™ 2 water system is a sealed system which does not allow for the rearranging of the cartridges. However, the final cartridge is an ORGANEX-Q® filter, followed by a final bacteria filter. Commercial laboratories using this system have not experienced any difficulty in using the water for culturing or testing. Reference to the MILLI-Q® systems throughout the remainder of the manual includes all MILLIPORE® or equivalent systems.

7.2.3 STANDARD, SYNTHETIC FRESHWATER

7.2.3.1 To prepare 20 L of synthetic, moderately hard, reconstituted water, use the reagent grade chemicals in Table 3 as follows:

1. Place 19 L of MILLI-Q®, or equivalent, water in a properly cleaned plastic carboy.
2. Add 1.20 g of $MgSO_4$, 1.92 g $NaHCO_3$, and 0.080g KCl to the carboy.
3. Aerate overnight.
4. Add 1.20 g of $CaSO_4 \cdot 2H_20$ to 1 L of MILLI-Q® or equivalent deionized water in a separate flask. Stir on magnetic stirrer until calcium sulfate is dissolved, add to the 19 L above, and mix well.
5. For *Ceriodaphnia dubia* culturing and testing, add sufficient sodium selenate (Na_2SeO_4) to provide 2 mg selenium per liter of final dilution water.
6. Aerate the combined solution vigorously for an additional 24 h to dissolve the added chemicals and stabilize the medium.
7. The measured pH, hardness, etc., should be as listed in Table 3.

TABLE 3. PREPARATION OF SYNTHETIC FRESHWATER USING REAGENT GRADE CHEMICALS[1]

Water Type	Reagent Added (mg/L)[2]				Approximate Final Water Quality		
	$NaHCO_3$	$CaSO_4 \cdot 2H_2O$	$MgSO_4$	KCl	pH[3]	Hardness[4]	Alkalinity[4]
Very soft	12.0	7.5	7.5	0.5	6.4-6.8	10-13	10-13
Soft	48.0	30.0	30.0	2.0	7.2-7.6	40-48	30-35
Moderately Hard	96.0	60.0	60.0	4.0	7.4-7.8	80-100	57-64
Hard	192.0	120.0	120.0	8.0	7.6-8.0	160-180	110-120
Very hard	384.0	240.0	240.0	16.0	8.0-8.4	280-320	225-245

1 Taken in part from Marking and Dawson (1973).
2 Add reagent grade chemicals to deionized water.
3 Approximate equilibrium pH after 24 h of aeration.
4 Expressed as mg $CaCO_3$/L.

7.2.3.2 If large volumes of synthetic reconstituted water will be needed, it may be advisable to mix 1 L portions of concentrated stock solutions of $NaHCO_3$, $MgSO_4$, and KCl for use in preparation of the reconstituted waters.

7.2.3.3 To prepare 20 L of standard, synthetic, moderately hard, reconstituted water, using mineral water such as PERRIER® Water, or equivalent (Table 4), follow the instructions below.

1. Place 16 L of MILLI-Q® or equivalent water in a properly cleaned plastic carboy.
2. Add 4 L of PERRIER® Water, or equivalent.
3. Aerate vigorously for 24 h to stabilize the medium.
4. The measured pH, hardness and alkalinity of the aerated water will be as indicated in Table 4.
5. This synthetic water is referred to as diluted mineral water (DMW) in the toxicity test methods.

TABLE 4. PREPARATION OF SYNTHETIC FRESHWATER USING MINERAL WATER[1]

Water Type	Volume of Mineral Water Added (mL/L)[2]	Proportion of Mineral Water (%)	Approximate Final Water Quality pH[3]	Hardness[4]	Alka-linity[4]
Very soft	50	2.5	7.2-8.1	10-13	10-13
Soft	100	10.0	7.9-8.3	40-48	30-35
Moderately Hard	200	20.0	7.9-8.3	80-100	57-64
Hard	400	40.0	7.9-8.3	160-180	110-120
Very hard[5]	---	---	---	---	---

[1] From Mount et al. (1987), and data provided by Philip Lewis, EMSL-Cincinnati, OH.
[2] Add mineral water to Milli-Q® water, or equivalent, to prepare Diluted Mineral Water (DMW).
[3] Approximate equilibrium pH after 24 h of aeration.
[4] Expressed as mg $CaCO_3$/L.
[5] Dilutions of PERRIER® Water form a precipitate when concentrations equivalent to "very hard water" are aerated.

7.3 USE OF RECEIVING WATER AS DILUTION WATER

7.3.1 If the objectives of the test require the use of uncontaminated receiving water as dilution water, and the receiving water is uncontaminated, it may be possible to collect a sample of the receiving water upstream of, or close to, but outside of the zone influenced by the effluent. However, if the receiving water is contaminated, it may be necessary to collect the sample in an area "remote" from the discharge site, matching as closely as possible the physical and chemical characteristics of the receiving water near the outfall.

7.3.2 The sample should be collected immediately prior to the test, but never more than 96 h before the test begins. Except where it is used within 24 h, or in the case where large volumes are required for flow through tests, the sample should be chilled to 0-6°C during or immediately following collection, and maintained at that temperature prior to use in the test.

7.3.3 Receiving water containing debris or indigenous organisms that may be confused with or attack the test organisms should be filtered through a sieve having 60 mm mesh openings prior to use.

7.3.4 Where toxicity-free dilution water is required in a test, the water is considered acceptable if test organisms show the required survival, growth, and reproduction in the controls during the test.

7.3.5 The regulatory authority may require that the hardness of the dilution water be comparable to the receiving water at the discharge site. This requirement can be satisfied by collecting an uncontaminated receiving water with a suitable hardness, or adjusting the hardness of an otherwise suitable receiving water by addition of reagents as indicated in Table 3.

7.4 **USE OF TAP WATER AS DILUTION WATER**

7.4.1 The use of tap water as dilution water is discouraged unless it is dechlorinated and passed through a deionizer and carbon filter. Tap water can be dechlorinated by deionization, carbon filtration, or the use of sodium thiosulfate. Use of 3.6 mg/L (anhydrous) sodium thiosulfate will reduce 1.0 mg chlorine/L (APHA, 1992). Following dechlorination, total residual chlorine should not exceed 0.01 mg/L. Because of the possible toxicity of thiosulfate to test organisms, a control lacking thiosulfate should be included in toxicity tests utilizing thiosulfate-dechlorinated water.

7.4.2 To be adequate for general laboratory use following dechlorination, the tap water is passed through a deionizer and carbon filter to remove toxic metals and organics, and to control hardness and alkalinity.

7.5 **DILUTION WATER HOLDING**

7.5.1 A given batch of dilution water should not be used for more than 14 days following preparation because of the possible build-up of bacterial, fungal, or algal slime growth and the problems associated with it. The container should be kept covered and the contents should be protected from light.

SECTION 8

EFFLUENT AND RECEIVING WATER SAMPLING, SAMPLE HANDLING, AND SAMPLE PREPARATION FOR TOXICITY TESTS

8.1 EFFLUENT SAMPLING

8.1.1 The effluent sampling point should be the same as that specified in the NPDES discharge permit (USEPA, 1988a). Conditions for exception would be: (l) better access to a sampling point between the final treatment and the discharge outfall; (2) if the processed waste is chlorinated prior to discharge, it may also be desirable to take samples prior to contact with the chlorine to determine toxicity of the unchlorinated effluent; or (3) in the event there is a desire to evaluate the toxicity of the influent to municipal waste treatment plants or separate wastewater streams in industrial facilities prior to their being combined with other wastewater streams or non-contact cooling water, additional sampling points may be chosen.

8.1.2 The decision on whether to collect grab or composite samples is based on the objectives of the test and an understanding of the short and long-term operations and schedules of the discharger. If the effluent quality varies considerably with time, which can occur where holding times are short, grab samples may seem preferable because of the ease of collection and the potential of observing peaks (spikes) in toxicity. However, the sampling duration of a grab sample is so short that full characterization of an effluent over a 24-h period would require a prohibitively large number of separate samples and tests. Collection of a 24-h composite sample, however, may dilute toxicity spikes, and average the quality of the effluent over the sampling period. Sampling recommendations are provided below (also see USEPA, 2002a).

8.1.3 Aeration during collection and transfer of effluents should be minimized to reduce the loss of volatile chemicals.

8.1.4 Details of date, time, location, duration, and procedures used for effluent sample and dilution water collection should be recorded.

8.2 EFFLUENT SAMPLE TYPES

8.2.1 The advantages and disadvantages of effluent grab and composite samples are listed below:

8.2.1.1 GRAB SAMPLES

Advantages:

1. Easy to collect; require a minimum of equipment and on-site time.
2. Provide a measure of instantaneous toxicity. Toxicity spikes are not masked by dilution.

Disadvantages:

1. Samples are collected over a very short period of time and on a relatively infrequent basis. The chances of detecting a spike in toxicity would depend on the frequency of sampling and the probability of missing a spike is high.

8.2.1.2 COMPOSITE SAMPLES

Advantages:

1. A single effluent sample is collected over a 24-h period.

2. The sample is collected over a much longer period of time than a single grab sample and contains all toxicity spikes.

Disadvantages:

1. Sampling equipment is more sophisticated and expensive, and must be placed on-site for at least 24 h.
2. Toxicity spikes may not be detected because they are masked by dilution with less toxic wastes.

8.3 EFFLUENT SAMPLING RECOMMENDATIONS

8.3.1 When tests are conducted on-site, test solutions can be renewed daily with freshly collected samples, except for the green alga, *Selenastrum capricornutum*, test which is not renewed.

8.3.2 When tests are conducted off-site, a minimum of three samples are collected. If these samples are collected on Test Days 1, 3, and 5, the first sample would be used for test initiation, and for test solution renewal on Day 2. The second sample would be used for test solution renewal on Days 3 and 4. The third sample would be used for test solution renewal on Days 5, 6, and 7.

8.3.3 Sufficient sample volume must be collected to perform the required toxicity and chemical tests. A 4-L (1-gal) CUBITAINER® will provide sufficient sample volume for most tests.

8.3.4 THE FOLLOWING EFFLUENT SAMPLING METHODS ARE RECOMMENDED:

8.3.4.1 Continuous Discharges

8.3.4.1.1 If the facility discharge is continuous, a single 24-h composite sample is to be taken.

8.3.4.2 Intermittent discharges

8.3.4.2.1 If the facility discharge is intermittent, a composite sample is to be collected for the duration of the discharge but not more than 24 hours.

8.4 RECEIVING WATER SAMPLING

8.4.1 Logistical problems and difficulty in securing sampling equipment generally preclude the collection of composite receiving water samples for toxicity tests. Therefore, based on the requirements of the test, a single grab sample or daily grab sample of receiving water is collected for use in the test.

8.4.2 The sampling point is determined by the objectives of the test. In rivers, samples should be collected from mid-stream and at mid-depth, if accessible. In lakes the samples are collected at mid-depth.

8.4.3 To determine the extent of the zone of toxicity in the receiving water downstream from the outfall, receiving water samples are collected at several distances downstream from the discharge. The time required for the effluent-receiving-water mixture to travel to sampling points downstream from the outfall, and the rate and degree of mixing, may be difficult to ascertain. Therefore, it may not be possible to correlate downstream toxicity with effluent toxicity at the discharge point unless a dye study is performed. The toxicity of receiving water samples from five stations downstream from the discharge point can be evaluated using the same number of test vessels and test organisms as used in one effluent toxicity test with five effluent dilutions.

8.5 **EFFLUENT AND RECEIVING WATER SAMPLE HANDLING, PRESERVATION, AND SHIPPING**

8.5.1 Unless the samples are used in an on-site toxicity test the day of collection (or hand delivered to the testing laboratory for use on the day of collection), they should be chilled and maintained at 0-6°C until used to inhibit microbial degradation, chemical transformations, and loss of highly volatile toxic substances.

8.5.2 Composite samples should be chilled as they are collected. Grab samples should be chilled immediately following collection.

8.5.3 If the effluent has been chlorinated, total residual chlorine must be measured immediately following sample collection.

8.5.4 Sample holding time begins when the last grab sample in a series is taken (i.e., when a series of four grab samples are taken over a 24-h period), or when a 24-h composite sampling period is completed. If the data from the samples are to be acceptable for use in the NPDES Program, the lapsed time (holding time) from sample collection to first use of each grab or composite sample must not exceed 36 h. EPA believes that 36 h is adequate time to deliver the samples to the laboratories performing the test in most cases. In the isolated cases, where the permittee can document that this delivery time cannot be met, the permitting authority can allow an option for on-site testing or a variance for an extension of shipped sample holding time. The request for a variance in sample holding time, directed to the USEPA Regional Administrator under 40 CFR 136.3(e) should include supportive data which show that the toxicity of the effluent sample is not reduced (e.g., because of volatilization and/or sorption of toxics on the sample container surfaces) by extending the holding time beyond more than 36 h. However, in no case should more than 72 h elapse between collection and first use of the sample. In static-renewal tests, each grab or composite sample may also be used to prepare test solutions for renewal at 24 h, 48 h, and/or 72 h after first use, if stored at 0-6°C, with minimum head space, as described in Subsection 8.5. If shipping problems (e.g., unsuccessful Saturday delivery) are encountered with renewal samples after a test has been initiated, the permitting authority may allow the continued use of the most recently used sample for test renewal. Guidance for determining the persistence of the sample is provided in Subsection 8.7.

8.5.5 To minimize the loss of toxicity due to volatilization of toxic constituents, all sample containers should be "completely" filled, leaving no air space between the contents and the lid.

8.5.6 SAMPLES USED IN ON-SITE TESTS

8.5.6.1 Samples collected for on-site tests should be used within 24 h.

8.5.7 SAMPLES SHIPPED TO OFF-SITE FACILITIES

8.5.7.1 Samples collected for off-site toxicity testing are to be chilled to 0-6°C during or immediately after collection, and shipped iced to the performing laboratory. Sufficient ice should be placed with the sample in the shipping container to ensure that ice will still be present when the sample arrives at the laboratory and is unpacked. Insulating material should not be placed between the ice and the sample in the shipping container unless required to prevent breakage of glass sample containers.

8.5.7.2 Samples may be shipped in one or more 4-L (l-gal) CUBITAINERS® or new plastic "milk" jugs. All sample containers should be rinsed with source water before being filled with sample. After use with receiving water or effluents, CUBITAINERS® and plastic jugs are punctured to prevent reuse.

8.5.7.3 Several sample shipping options are available, including Express Mail, air express, bus, and courier service. Express Mail is delivered seven days a week. Saturday and Sunday shipping and receiving schedules of private carriers vary with the carrier.

8.6 SAMPLE RECEIVING

8.6.1 Upon arrival at the laboratory, samples are logged in and the temperature is measured and recorded. If the samples are not immediately prepared for testing, they are stored at 0-6°C until used.

8.6.2 Every effort must be made to initiate the test with an effluent sample on the day of arrival in the laboratory, and the sample holding time should not exceed 36 h unless a variance has been granted by the NPDES permitting authority.

8.7 PERSISTENCE OF EFFLUENT TOXICITY DURING SAMPLE SHIPMENT AND HOLDING

8.7.1 The persistence of the toxicity of an effluent prior to its use in a toxicity test is of interest in assessing the validity of toxicity test data, and in determining the possible effects of allowing an extension of the holding time. Where a variance in holding time (> 36 h, but ≤ 72 h) is requested by a permittee, (see Subsection 8.5.4 above), information on the effects of the extension in holding time on the toxicity of samples must be obtained by comparing the results of multi-concentration chronic toxicity tests performed on effluent samples held 36 h with toxicity test results using the same samples after they were held for the requested, longer period. The portion of the sample set aside for the second test should be held under the same conditions as during shipment and holding.

8.8 PREPARATION OF EFFLUENT AND RECEIVING WATER SAMPLES FOR TOXICITY TESTS

8.8.1 When aliquots are removed from the sample container, the head space above the remaining sample should be held to a minimum. Air which enters a container upon removal of sample should be expelled by compressing the container before reclosing, if possible (i.e., where a CUBITAINER® is used), or by using an appropriate discharge valve (spigot).

8.8.2 With the daphnid, *Ceriodaphnia dubia*, and fathead minnow, *Pimephales promelas*, tests, effluents and receiving waters should be filtered through a 60-µm plankton net to remove indigenous organisms that may attack or be confused with the test organisms (see the daphnid, *Ceriodaphnia dubia*, test method for details). Receiving waters used in green alga, *Selenastrum capricornutum*, toxicity tests must be filtered through a 0.45-µm pore diameter filter before use. It may be necessary to first coarse-filter the dilution and/or waste water through a nylon sieve having 2- to 4-mm mesh openings to remove debris and/or break up large floating or suspended solids. Because filtration may increase the dissolved oxygen (DO) in the effluent, the DO should be checked both before and after filtering. Low dissolved oxygen concentrations will indicate a potential problem in performing the test. **Caution:** filtration may remove some toxicity.

8.8.3 If the samples must be warmed to bring them to the prescribed test temperature, supersaturation of the dissolved oxygen and nitrogen may become a problem. To avoid this problem, samples may be warmed slowly in open test containers. If DO is still above 100% saturation after warming to test temperature, samples should be aerated moderately (approximately 500 mL/min) for a few minutes using an airstone. If DO is below 4.0 mg/L after warming to test temperature, the solutions must be aerated moderately (approximately 500 mL/min) for a few minutes, using an airstone, until the DO is within the prescribed range (≥4.0 mg/L). **Caution:** avoid excessive aeration.

8.8.4 The DO concentration in the samples should be near saturation prior to use. Aeration may be used to bring the DO and other gases into equilibrium with air, minimize oxygen demand, and stabilize the pH. However, aeration during collection, transfer, and preparation of samples should be minimized to reduce the loss of volatile chemicals.

8.8.4.1 Aeration during the test may alter the results and should be used only as a last resort to maintain the required DO. Aeration can reduce the apparent toxicity of the test solutions by stripping them of highly volatile toxic substances, or increase their toxicity by altering pH. However, the DO in the test solutions should not be allowed to fall below 4.0 mg/L.

8.8.4.2 In static tests (renewal or non-renewal), low DOs may commonly occur in the higher concentrations of wastewater. Aeration is accomplished by bubbling air through a pipet at a rate of 100 bubbles/min. If aeration is necessary, all test solutions must be aerated. It is advisable to monitor the DO closely during the first few hours of the test. Samples with a potential DO problem generally show a downward trend in DO within 4 to 8 h after the test is started. Unless aeration is initiated during the first 8 h of the test, the DO may be exhausted during an unattended period, thereby invalidating the test.

8.8.5 At a minimum, pH, conductivity, and total residual chlorine are measured in the undiluted effluent or receiving water, and pH and conductivity are measured in the dilution water.

8.8.5.1 It is recommended that total alkalinity and total hardness also be measured in the undiluted effluent test water, receiving water, and the dilution water.

8.8.6 Total ammonia is measured in effluent and receiving water samples where toxicity may be contributed by un-ionized ammonia (i.e., where total ammonia $\geq$ 5 mg/L). The concentration (mg/L) of un-ionized (free) ammonia in a sample is a function of temperature and pH, and is calculated using the percentage value obtained from Table 5, under the appropriate pH and temperature, and multiplying it by the concentration (mg/L) of total ammonia in the sample.

8.8.7 Effluents and receiving waters can be dechlorinated using 6.7 mg/L anhydrous sodium thiosulfate to reduce 1 mg/L chlorine (APHA, 1992). Note that the amount of thiosulfate required to dechlorinate effluents is greater than the amount needed to dechlorinate tap water (see Section 7, Dilution Water, Subsection 7.4.1). Since thiosulfate may contribute to sample toxicity, a thiosulfate control should be used in the test in addition to the normal dilution water control.

8.8.8 Mortality or impairment of growth or reproduction due to pH alone may occur if the pH of the sample falls outside the range of 6.0 - 9.0. Thus, the presence of other forms of toxicity (metals and organics) in the sample may be masked by the toxic effects of low or high pH. The question about the presence of other toxicants can be answered only by performing two parallel tests, one with an adjusted pH, and one without an adjusted pH. Freshwater samples are adjusted to pH 7.0 by adding 1N NaOH or 1N HCl dropwise, as required, being careful to avoid overadjustment.

TABLE 5. PERCENT UNIONIZED NH_3 IN AQUEOUS AMMONIA SOLUTIONS: TEMPERATURES 15-26°C AND pH 6.0-8.9[1]

pH	TEMPERATURE (°C)											
	15	16	17	18	19	20	21	22	23	24	25	26
6.0	0.0274	0.0295	0.0318	0.0343	0.0369	0.0397	0.0427	0.0459	0.0493	0.0530	0.0568	0.0610
6.1	0.0345	0.0372	0.0400	0.0431	0.0464	0.0500	0.0537	0.0578	0.0621	0.0667	0.0716	0.0768
6.2	0.0434	0.0468	0.0504	0.0543	0.0584	0.0629	0.0676	0.0727	0.0781	0.0901	0.0901	0.0966
6.3	0.0546	0.0589	0.0634	0.0683	0.0736	0.0792	0.0851	0.0915	0.0983	0.1134	0.1134	0.1216
6.4	0.0687	0.0741	0.0799	0.0860	0.0926	0.0996	0.107	0.115	0.124	0.133	0.143	0.153
6.5	0.0865	0.0933	0.1005	0.1083	0.1166	0.1254	0.135	0.145	0.156	0.167	0.180	0.193
6.6	0.109	0.117	0.127	0.136	0.147	0.158	0.170	0.182	0.196	0.210	0.226	0.242
6.7	0.137	0.148	0.159	0.171	0.185	0.199	0.214	0.230	0.247	0.265	0.284	0.305
6.8	0.172	0.186	0.200	0.216	0.232	0.250	0.269	0.289	0.310	0.333	0.358	0.384
6.9	0.217	0.234	0.252	0.271	0.292	0.314	0.338	0.363	0.390	0.419	0.450	0.482
7.0	0.273	0.294	0.317	0.342	0.368	0.396	0.425	0.457	0.491	0.527	0.566	0.607
7.1	0.343	0.370	0.399	0.430	0.462	0.497	0.535	0.575	0.617	0.663	0.711	0.762
7.2	0.432	0.466	0.502	0.540	0.581	0.625	0.672	0.722	0.776	0.833	0.893	0.958
7.3	0.543	0.586	0.631	0.679	0.731	0.786	0.845	0.908	0.975	1.05	1.12	1.20
7.4	0.683	0.736	0.793	0.854	0.918	0.988	1.061	1.140	1.224	1.31	1.41	1.51
7.5	0.858	0.925	0.996	1.07	1.15	1.24	1.33	1.43	1.54	1.65	1.77	1.89
7.6	1.08	1.16	1.25	1.35	1.45	1.56	1.67	1.80	1.93	2.07	2.21	2.37
7.7	1.35	1.46	1.57	1.69	1.82	1.95	2.10	2.25	2.41	2.59	2.77	2.97
7.8	1.70	1.83	1.97	2.12	2.28	2.44	2.62	2.82	3.02	3.24	3.46	3.71
7.9	2.13	2.29	2.46	2.65	2.85	3.06	3.28	3.52	3.77	4.04	4.32	4.62
8.0	2.66	2.87	3.08	3.31	3.56	3.82	4.10	4.39	4.70	5.03	5.38	5.75
8.1	3.33	3.58	3.85	4.14	4.44	4.76	5.10	5.46	5.85	6.25	6.68	7.14
8.2	4.16	4.47	4.80	5.15	5.52	5.92	6.34	6.78	7.25	7.75	8.27	8.82
8.3	5.18	5.56	5.97	6.40	6.86	7.34	7.85	8.39	8.96	9.56	10.2	10.9
8.4	6.43	6.90	7.40	7.93	8.48	9.07	9.69	10.3	11.0	11.7	12.5	13.3
8.5	7.97	8.54	9.14	9.78	10.45	11.16	11.90	12.7	13.5	14.4	15.2	16.2
8.6	9.83	10.5	11.2	12.0	12.8	13.6	14.5	15.5	16.4	17.4	18.5	19.5
8.7	12.07	12.9	13.8	14.7	15.6	16.6	17.6	18.7	19.8	21.0	22.2	23.4
8.8	14.7	15.7	16.7	17.8	18.9	20.0	21.2	22.5	23.7	25.1	26.4	27.8
8.9	17.9	19.0	20.2	21.4	22.7	24.0	25.3	26.7	28.2	29.6	31.1	32.6

[1] Table provided by Teresa Norberg-King, ERL, Duluth, Minnesota. Also see Emerson et al. (1975), Thurston et al. (1974), and USEPA (1985a).

8.9 **PRELIMINARY TOXICITY RANGE-FINDING TESTS**

8.9.1 USEPA Regional and State personnel generally have observed that it is not necessary to conduct a toxicity range-finding test prior to initiating a static, chronic, definitive toxicity test. However, when preparing to perform a static test with a sample of completely unknown quality, or before initiating a flow-through test, it is advisable to conduct a preliminary toxicity range-finding test.

8.9.2 A toxicity range-finding test ordinarily consists of a down-scaled, abbreviated static acute test in which groups of five organisms are exposed to several widely-spaced sample dilutions in a logarithmic series, such as

100%, 10.0%, 1.00%, and 0.100%, and a control, for 8-24 h. **Caution**: if the sample must also be used for the full-scale definitive test, the 36-h limit on holding time (see Subsection 8.5.4) must not be exceeded before the definitive test is initiated.

8.9.3 It should be noted that the toxicity (LC50) of a sample observed in a range-finding test may be significantly different from the toxicity observed in the follow-up chronic definitive test because: (1) the definitive test is longer; and (2) the test may be performed with a sample collected at a different time, and possibly differing significantly in the level of toxicity.

8.10 MULTI-CONCENTRATION (DEFINITIVE) EFFLUENT TOXICITY TESTS

8.10.1 The tests recommended for use in determining discharge permit compliance in the NPDES program are multi-concentration, or definitive, tests which provide (1) a point estimate of effluent toxicity in terms of an IC25, IC50, or LC50, or (2) a no-observed-effect-concentration (NOEC) defined in terms of mortality, growth, reproduction, and/or teratogenicity and obtained by hypothesis testing. The tests may be static renewal or static non-renewal.

8.10.2 The tests consist of a control and a minimum of five effluent concentrations. USEPA recommends the use of a $\geq$0.5 dilution factor for selecting effluent test concentrations. Effluent test concentrations of 6.25%, 12.5%, 25%, 50%, and 100% are commonly used, however, test concentrations should be selected independently for each test based on the objective of the study, the expected range of toxicity, the receiving water concentration, and any available historical testing information on the effluent. USEPA (2000a) provides additional guidance on choosing appropriate test concentrations.

8.10.3 When these tests are used in determining compliance with permit limits, effluent test concentrations should be selected to bracket the receiving water concentration. This may be achieved by selecting effluent test concentrations in the following manner: (1) 100% effluent, (2) [RWC + 100]/2, (3) RWC, (4) RWC/2, and (5) RWC/4. For example, where the RWC = 50%, appropriate effluent concentrations may be 100%, 75%, 50%, 25%, and 12.5%.

8.10.4 If acute/chronic ratios are to be determined by simultaneous acute and short-term chronic tests with a single species, using the same sample, both types of tests must use the same test conditions, i.e., pH, temperature, water hardness, salinity, etc.

8.11 RECEIVING WATER TESTS

8.11.1 Receiving water toxicity tests generally consist of 100% receiving water and a control. The total hardness of the control should be comparable to the receiving water.

8.11.2 The data from the two treatments are analyzed by hypothesis testing to determine if test organism survival in the receiving water differs significantly from the control. Four replicates and 10 organisms per replicate are required for each treatment (see Summary of Test Conditions and Test Acceptability Criteria in the specific test method).

8.11.3 In cases where the objective of the test is to estimate the degree of toxicity of the receiving water, a multi-concentration test is performed by preparing dilutions of the receiving water, using a $\geq$ 0.5 dilution series, with a suitable control water.

SECTION 9

CHRONIC TOXICITY TEST ENDPOINTS AND DATA ANALYSIS

9.1 ENDPOINTS

9.1.1 The objective of chronic aquatic toxicity tests with effluents and pure compounds is to estimate the highest "safe" or "no-effect concentration" of these substances. For practical reasons, the responses observed in these tests are usually limited to hatchability, gross morphological abnormalities, survival, growth, and reproduction, and the results of the tests are usually expressed in terms of the highest toxicant concentration that has no statistically significant observed effect on these responses, when compared to the controls. The terms currently used to define the endpoints employed in the rapid, chronic and sub-chronic toxicity tests have been derived from the terms previously used for full life-cycle tests. As shorter chronic tests were developed, it became common practice to apply the same terminology to the endpoints. The terms used in this manual are as follows:

9.1.1.1 Safe Concentration - The highest concentration of toxicant that will permit normal propagation of fish and other aquatic life in receiving waters. The concept of a "safe concentration" is a biological concept, whereas the "no-observed-effect concentration" (below) is a statistically defined concentration.

9.1.1.2 No-Observed-Effect-Concentration (NOEC) - The highest concentration of toxicant to which organisms are exposed in a full life-cycle or partial life-cycle (short-term) test, that causes no observable adverse effects on the test organisms (i.e., the highest concentration of toxicant in which the values for the observed responses are not statistically significantly different from the controls). This value is used, along with other factors, to determine toxicity limits in permits.

9.1.1.3 Lowest-Observed-Effect-Concentration (LOEC) - The lowest concentration of toxicant to which organisms are exposed in a life-cycle or partial life-cycle (short-term) test, which causes adverse effects on the test organisms (i.e., where the values for the observed responses are statistically significantly different from the controls).

9.1.1.4 Effective Concentration (EC) - A point estimate of the toxicant concentration that would cause an observable adverse affect on a quantal, "all or nothing," response (such as death, immobilization, or serious incapacitation) in a given percent of the organisms, calculated by point estimation techniques. If the observable effect is death or immobility, the term, Lethal Concentration (LC), should be used (see Subsection 9.1.1.5). A certain EC or LC value might be judged from a biological standpoint to represent a threshold concentration, or lowest concentration that would cause an adverse effect on the observed response.

9.1.1.5 Lethal Concentration (LC) - The toxicant concentration that would cause death in a given percent of the test population. Identical to EC when the observed adverse effect is death. For example, the LC50 is the concentration of toxicant that would cause death in 50% of the test population.

9.1.1.6 Inhibition Concentration (IC) - The toxicant concentration that would cause a given percent reduction in a non-quantal biological measurement for the test population. For example, the IC25 is the concentration of toxicant that would cause a 25% reduction in mean young per female or in growth for the test population, and the IC50 is the concentration of toxicant that would cause a 50% reduction.

9.2 RELATIONSHIP BETWEEN ENDPOINTS DETERMINED BY HYPOTHESIS TESTING AND POINT ESTIMATION TECHNIQUES

9.2.1 If the objective of chronic aquatic toxicity tests with effluents and pure compounds is to estimate the highest "safe or no-effect concentration" of these substances, it is imperative to understand how the statistical endpoints of these tests are related to the "safe" or "no-effect" concentration. NOECs and LOECs are determined by hypothesis testing (Dunnett's Test, a t test with the Bonferroni adjustment, Steel's Many-one Rank Test, or the Wilcoxon Rank

Sum Test with the Bonferroni adjustment), whereas LCs, ICs, and ECs are determined by point estimation techniques (Probit Analysis, Spearman-Karber Method, Trimmed Spearman-Karber Method, Graphical Method or Linear Interpolation Method). There are inherent differences between the use of a NOEC or LOEC derived from hypothesis testing to estimate a "safe" concentration, and the use of a LC, EC, IC, or other point estimates derived from curve fitting, interpolation, etc.

9.2.2 Most point estimates, such as the LC, IC, or EC, are derived from a mathematical model that assumes a continuous dose-response relationship. By definition, any LC, IC, or EC value is an estimate of some amount of adverse effect. Thus the assessment of a "safe" concentration must be made from a biological standpoint rather than with a statistical test. In this instance, the biologist must determine some amount of adverse effect that is deemed to be "safe", in the sense that from a practical biological viewpoint it will not affect the normal propagation of fish and other aquatic life in receiving waters.

9.2.3 The use of NOECs and LOECs, on the other hand, assumes either (1) a continuous dose-response relationship, or (2) a non-continuous (threshold) model of the dose-response relationship.

9.2.3.1 In the case of a continuous dose-response relationship, it is also assumed that adverse effects that are not "statistically observable" are also not important from a biological standpoint, since they are not pronounced enough to test as statistically significant against some measure of the natural variability of the responses.

9.2.3.2 In the case of non-continuous dose-response relationships, it is assumed that there exists a true threshold, or concentration below which there is no adverse effect on aquatic life, and above which there is an adverse effect. The purpose of the statistical analysis in this case is to estimate as closely as possible where that threshold lies.

9.2.3.3 In either case, it is important to realize that the amount of adverse effect that is statistically observable (LOEC) or not observable (NOEC) is highly dependent on all aspects of the experimental design, such as the number of concentrations of toxicant, number of replicates per concentration, number of organisms per replicate, and use of randomization. Other factors that affect the sensitivity of the test include the choice of statistical analysis, the choice of an alpha level, and the amount of variability between responses at a given concentration.

9.2.3.4 Where the assumption of a continuous dose-response relationship is made, by definition some amount of adverse effect might be present at the NOEC, but is not great enough to be detected by hypothesis testing.

9.2.3.5 Where the assumption of a non-continuous dose-response relationship is made, the NOEC would indeed be an estimate of a "safe" or "no-effect" concentration if the amount of adverse effect that appears at the threshold is great enough to test as statistically significantly different from the controls in the face of all aspects of the experimental design mentioned above. If, however, the amount of adverse effect at the threshold were not great enough to test as statistically different, some amount of adverse effect might be present at the NOEC. In any case, the estimate of the NOEC with hypothesis testing is always dependent on the aspects of the experimental design mentioned above. For this reason, the reporting and examination of some measure of the sensitivity of the test (either the minimum significant difference or the percent change from the control that this minimum difference represents) is extremely important.

9.2.4 In summary, the assessment of a "safe" or "no-effect" concentration cannot be made from the results of statistical analysis alone, unless (1) the assumptions of a strict threshold model are accepted, and (2) it is assumed that the amount of adverse effect present at the threshold is statistically detectable by hypothesis testing. In this case, estimates obtained from a statistical analysis are indeed estimates of a "no-effect" concentration. If the assumptions are not deemed tenable, then estimates from a statistical analysis can only be used in conjunction with an assessment from a biological standpoint of what magnitude of adverse effect constitutes a "safe" concentration. In this instance, a "safe" concentration is not necessarily a truly "no-effect" concentration, but rather a concentration at which the effects are judged to be of no biological significance.

9.2.5 A better understanding of the relationship between endpoints derived by hypothesis testing (NOECs) and point estimation techniques (LCs, ICs, and ECs) would be very helpful in choosing methods of data analysis. Norberg-King (1991) reported that the IC25s were comparable to the NOECs for 23 effluent and reference toxicant data sets analyzed. The data sets included short-term chronic toxicity tests for the fathead minnow, *Pimephales promelas*, and the daphnid, *Ceriodaphnia dubia*. Birge et al. (1985) reported that LC1s derived from Probit Analysis of data from short-term embryo-larval tests with reference toxicants were comparable to NOECs for several organisms. Similarly, USEPA (1988d) reported that the IC25s were comparable to the NOECs for a set of daphnid, *Ceriodaphnia dubia*, chronic tests with a single reference toxicant. However, the scope of these comparisons was very limited, and sufficient information is not yet available to establish an overall relationship between these two types of endpoints, especially when derived from effluent toxicity test data.

9.3 **PRECISION**

9.3.1 HYPOTHESIS TESTS

9.3.1.1 When hypothesis tests are used to analyze toxicity test data, it is not possible to express precision in terms of a commonly used statistic. The results of the test are given in terms of two endpoints, the No-Observed- Effect Concentration (NOEC) and the Lowest-Observed-Effect Concentration (LOEC). The NOEC and LOEC are limited to the concentrations selected for the test. The width of the NOEC-LOEC interval is a function of the dilution series, and differs greatly depending on whether a dilution factor of 0.3 or 0.5 is used in the test design. Therefore, **USEPA recommends the use of the ≥ 0.5 dilution factor** (see Section 4, Quality Assurance). It is not possible to place confidence limits on the NOEC and LOEC derived from a given test, and it is difficult to quantify the precision of the NOEC-LOEC endpoints between tests. If the data from a series of tests performed with the same toxicant, toxicant concentrations, and test species, were analyzed with hypothesis tests, precision could only be assessed by a qualitative comparison of the NOEC-LOEC intervals, with the understanding that maximum precision would be attained if all tests yielded the same NOEC-LOEC interval. In practice, the precision of results of repetitive chronic tests is considered acceptable if the NOECs vary by no more than one concentration interval above or below a central tendency. Using these guidelines, the "normal" range of NOECs from toxicity tests using a 0.5 dilution factor (two-fold difference between adjacent concentrations), would be four-fold.

9.3.2 POINT ESTIMATION TECHNIQUES

9.3.2.1 Point estimation techniques have the advantage of providing a point estimate of the toxicant concentration causing a given amount of adverse (inhibiting) effect, the precision of which can be quantitatively assessed (1) within tests by calculation of 95% confidence limits, and (2) across tests by calculating a standard deviation and coefficient of variation.

9.3.2.2 It should be noted that software used to calculate point estimates occasionally may not provide associated 95% confidence intervals. This situation may arise when test data do not meet specific assumptions required by the statistical methods, when point estimates are outside of the test concentration range, and when specific limitations imposed by the software are encountered. USEPA (2000a) provides guidance on confidence intervals under these circumstances.

9.4 **DATA ANALYSIS**

9.4.1 ROLE OF THE STATISTICIAN

9.4.1.1 The use of the statistical methods described in this manual for routine data analysis does not require the assistance of a statistician. However, the interpretation of the results of the analysis of the data from any of the toxicity tests described in this manual can become problematic because of the inherent variability and sometimes unavoidable anomalies in biological data. If the data appear unusual in any way, or fail to meet the necessary assumptions, a statistician should be consulted. Analysts who are not proficient in statistics are strongly advised to seek the assistance of a statistician before selecting the method of analysis and using any of the results.

9.4.1.2 The statistical methods recommended in this manual are not the only possible methods of statistical analysis. Many other methods have been proposed and considered. Certainly there are other reasonable and defensible methods of statistical analysis for this kind of toxicity data. Among alternative hypothesis tests some, like Williams' Test, require additional assumptions, while others, like the bootstrap methods, require computer-intensive computations. Alternative point estimation approaches most probably would require the services of a statistician to determine the appropriateness of the model (goodness of fit), higher order linear or nonlinear models, confidence intervals for estimates generated by inverse regression, etc. In addition, point estimation or regression approaches would require the specification by biologists or toxicologists of some low level of adverse effect that would be deemed acceptable or safe. The statistical methods contained in this manual have been chosen because they are (1) applicable to most of the different toxicity test data sets for which they are recommended, (2) powerful statistical tests, (3) hopefully "easily" understood by nonstatisticians, and (4) amenable to use without a computer, if necessary.

9.4.2 PLOTTING THE DATA

9.4.2.1 The data should be plotted, both as a preliminary step to help detect problems and unsuspected trends or patterns in the responses, and as an aid in interpretation of the results. Further discussion and plotted sets of data are included in the methods and the Appendices.

9.4.3 DATA TRANSFORMATIONS

9.4.3.1 Transformations of the data, (e.g., arc sine square root and logs), are used where necessary to meet assumptions of the proposed analyses, such as the requirement for normally distributed data.

9.4.4 INDEPENDENCE, RANDOMIZATION, AND OUTLIERS

9.4.4.1 Statistical independence among observations is a critical assumption in all statistical analysis of toxicity data. One of the best ways to insure independence is to properly follow rigorous randomization procedures. Randomization techniques should be employed at the start of the test, including the randomization of the placement of test organisms in the test chambers and randomization of the test chamber location within the array of chambers. Discussions of statistical independence, outliers and randomization, and a sample randomization scheme, are included in Appendix A.

9.4.5 REPLICATION AND SENSITIVITY

9.4.5.1 The number of replicates employed for each toxicant concentration is an important factor in determining the sensitivity of chronic toxicity tests. Test sensitivity generally increases as the number of replicates is increased, but the point of diminishing returns in sensitivity may be reached rather quickly. The level of sensitivity required by a hypothesis test or the confidence interval for a point estimate will determine the number of replicates, and should be based on the objectives for obtaining the toxicity data.

9.4.5.2 In a statistical analysis of toxicity data, the choice of a particular analysis and the ability to detect departures from the assumptions of the analysis, such as the normal distribution of the data and homogeneity of variance, is also dependent on the number of replicates. More than the minimum number of replicates may be required in situations where it is imperative to obtain optimal statistical results, such as with tests used in enforcement cases or when it is not possible to repeat the tests. For example, when the data are analyzed by hypothesis testing, the nonparametric alternatives cannot be used unless there are at least four replicates at each toxicant concentration.

9.4.6 RECOMMENDED ALPHA LEVELS

9.4.6.1 The data analysis examples included in the manual specify an alpha level of 0.01 for testing the assumptions of hypothesis tests and an alpha level of 0.05 for the hypothesis tests themselves. These levels are

common and well accepted levels for this type of analysis and are presented as a recommended minimum significance level for toxicity test data analysis.

9.5 CHOICE OF ANALYSIS

9.5.1 The recommended statistical analysis of most data from chronic toxicity tests with aquatic organisms follows a decision process illustrated in the flowchart in Figure 2. An initial decision is made to use point estimation techniques (the Probit Analysis, the Spearman-Karber Method, the Trimmed Spearman-Karber Method, the Graphical Method, or Linear Interpolation Method) and/or to use hypothesis testing (Dunnett's Test, the t test with the Bonferroni adjustment, Steel's Many-one Rank Test, or the Wilcoxon Rank Sum Test with the Bonferroni adjustment). **NOTE: For the NPDES Permit Program, the point estimation techniques are the preferred statistical methods in calculating end points for effluent toxicity tests.** If hypothesis testing is chosen, subsequent decisions are made on the appropriate procedure for a given set of data, depending on the results of the tests of assumptions, as illustrated in the flowchart. A specific flow chart is included in the analysis section for each test.

9.5.2 Since a single chronic toxicity test might yield information on more than one parameter (such as survival, growth, and reproduction), the lowest estimate of a "no-observed-effect concentration" for any of the responses would be used as the "no-observed-effect concentration" for each test. It follows logically that in the statistical analysis of the data, concentrations that had a significant toxic effect on one of the observed responses would not be subsequently tested for an effect on some other response. This is one reason for excluding concentrations that have shown a statistically significant reduction in survival from a subsequent hypothesis test for effects on another parameter such as reproduction. A second reason is that the exclusion of such concentrations usually results in a more powerful and appropriate statistical analysis. In performing the point estimation techniques recommended in this manual, an all-data approach is used. For example, data from concentrations above the NOEC for survival are included in determining ICp estimates using the Linear Interpolation Method.

9.5.3 ANALYSIS OF GROWTH AND REPRODUCTION DATA

9.5.3.1 Growth data from the fathead minnow, *Pimephales promelas*, larval survival and growth test are analyzed using hypothesis testing or point estimation techniques according to the flowchart in Figure 2. The above mentioned growth data may also be analyzed by generating a point estimate with the Linear Interpolation Method. Data from effluent concentrations that have tested significantly different from the control for survival are excluded from further hypothesis tests concerning growth effects. Growth is defined as the dry weight per original number of test organisms when group weights are obtained. When analyzing the data using point estimation techniques, data from all concentrations are included in the analysis.

9.5.3.2 Reproduction data from the daphnid, *Ceriodaphnia dubia*, survival and reproduction test are analyzed using hypothesis testing or point estimation techniques according to the flowchart in Figure 2. In hypothesis testing, data from effluent concentrations that have significantly lower survival than the control, as determined by Fisher's Exact test, are not included in the hypothesis tests for reproductive effects. Data from all concentrations are included when using point estimation techniques.

9.5.4 ANALYSIS OF ALGAL GROWTH RESPONSE DATA

9.5.4.1 The growth response data from the green alga, *Selenastrum capricornutum*, toxicity test, after an appropriate transformation, if necessary, to meet the assumptions of normality and homogeneity of variance, may be analyzed by hypothesis testing according to the flowchart in Figure 2. Point estimates, such as the IC25 and IC50, would also be appropriate in analyzing algal growth data.

9.5.5 ANALYSIS OF MORTALITY DATA

9.5.5.1 Mortality data are analyzed by Probit Analysis, if appropriate, or other point estimation techniques (i.e., the Spearman-Karber Method, the Trimmed Spearman-Karber Method, or the Graphical Method) (see Appendices I-L and the discussion below). The mortality data can also be analyzed by hypothesis testing, after an arc sine square root transformation (see Appendix B-F), according to the flowchart in Figure 2.

9.5.5.2 Mortality data from the daphnid, *Ceriodaphnia dubia*, survival and reproduction test are analyzed by Fisher's Exact Test (Appendix G) prior to the analysis of the reproduction data. The mortality data may also be analyzed by Probit Analysis, if appropriate or other methods (see Subsection 9.5.5.1).

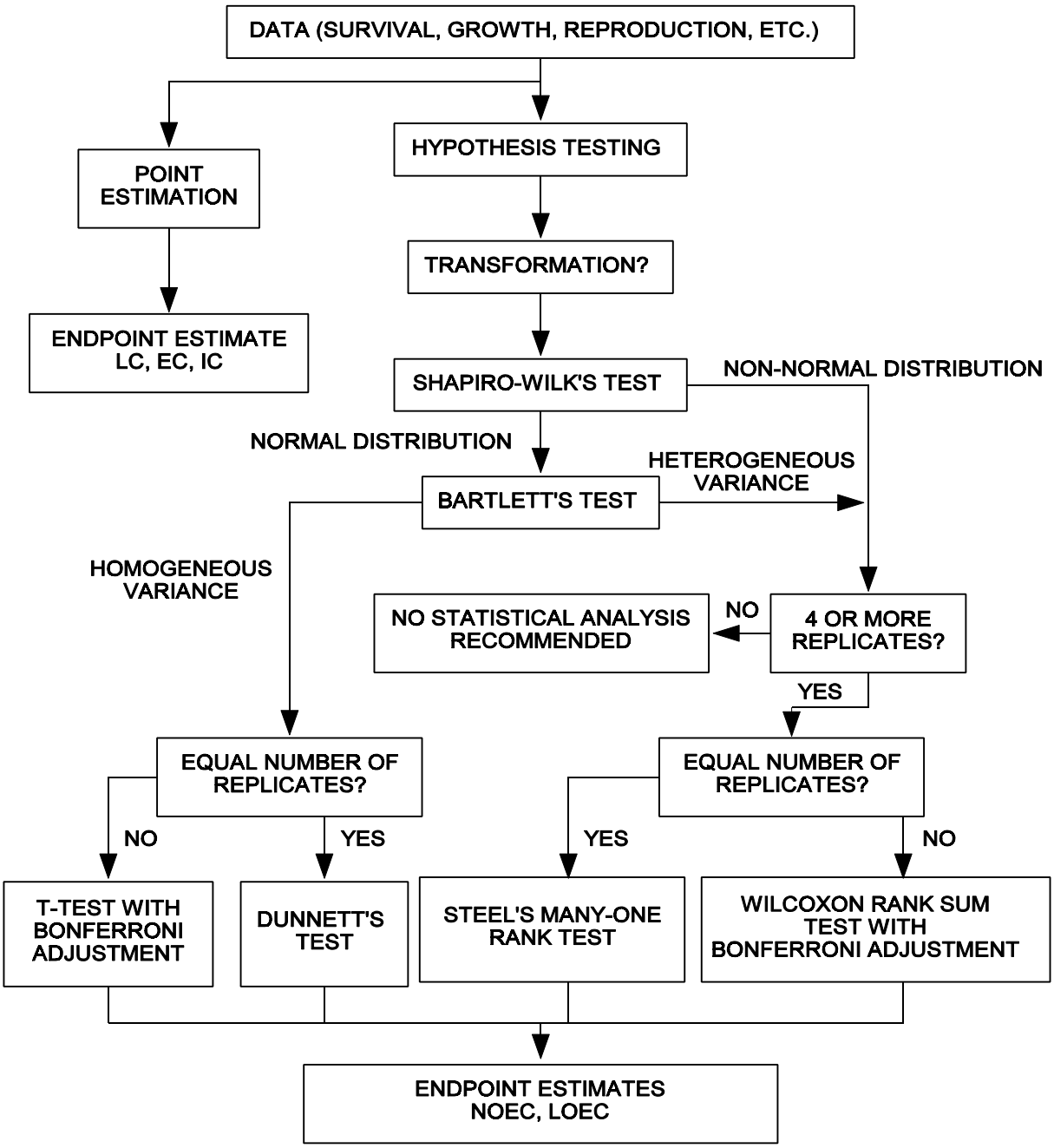

Figure 2. Flowchart for statistical analysis of test data

9.6 **HYPOTHESIS TESTS**

9.6.1 DUNNETT'S PROCEDURE

9.6.1.1 Dunnett's Procedure is used to determine the NOEC. The procedure consists of an analysis of variance (ANOVA) to determine the error term, which is then used in a multiple comparison procedure for comparing each of the treatment means with the control mean, in a series of paired tests (see Appendix C). Use of Dunnett's Procedure requires at least three replicates per treatment to check the assumptions of the test. In cases where the numbers of data points (replicates) for each concentration are not equal, a t test may be performed with Bonferroni's adjustment for multiple comparisons (see Appendix D), instead of using Dunnett's Procedure.

9.6.1.2 The assumptions upon which the use of Dunnett's Procedure is contingent are that the observations within treatments are normally distributed, with homogeneity of variance. Before analyzing the data, these assumptions must be tested using the procedures provided in Appendix B.

9.6.1.3 If, after suitable transformations have been carried out, the normality assumptions have not been met, Steel's Many-one Rank Test should be used if there are four or more data points (replicates) per toxicant concentration. If the numbers of data points for each toxicant concentration are not equal, the Wilcoxon Rank Sum Test with Bonferroni's adjustment should be used (see Appendix F).

9.6.1.4 Some indication of the sensitivity of the analysis should be provided by calculating (1) the minimum difference between means that can be detected as statistically significant, and (2) the percent change from the control mean that this minimum difference represents for a given test.

9.6.1.5 A step-by-step example of the use of Dunnett's Procedure is provided in Appendix C.

9.6.2 T TEST WITH THE BONFERRONI ADJUSTMENT

9.6.2.1 A t test with Bonferroni's adjustment is used as an alternative to Dunnett's Procedure when the number of replicates is not the same for all concentrations. This test sets an upper bound of alpha on the overall error rate, in contrast to Dunnett's Procedure, for which the overall error rate is fixed at alpha. Thus Dunnett's Procedure is a more powerful test.

9.6.2.2 The assumptions upon which the use of the t test with Bonferroni's adjustment is contingent are that the observations within treatments are normally distributed, with homogeneity of variance. These assumptions must be tested using the procedures provided in Appendix B.

9.6.2.3 The estimate of the safe concentration derived from this test is reported in terms of the NOEC. A step-by-step example of the use of the t test with Bonferroni's adjustment is provided in Appendix D.

9.6.3 STEEL'S MANY-ONE RANK TEST

9.6.3.1 Steel's Many-one Rank Test is a multiple comparison procedure for comparing several treatments with a control. This method is similar to Dunnett's Procedure, except that it is not necessary to meet the assumption of normality. The data are ranked, and the analysis is performed on the ranks rather than on the data themselves. If the data are normally or nearly normally distributed, Dunnett's Procedure would be more sensitive (would detect smaller differences between the treatments and control). For data that are not normally distributed, Steel's Many-one Rank Test can be much more efficient (Hodges and Lehmann, 1956).

9.6.3.2 It is necessary to have at least four replicates per toxicant concentration to use Steel's test. Unlike Dunnett's procedure, the sensitivity of this test cannot be stated in terms of the minimum difference between treatment means and the control mean that can be detected as statistically significant.

9.6.3.3 The estimate of the safe concentration is reported as the NOEC. A step-by-step example of the use of Steel's Many-one Rank Test is provided in Appendix E.

9.6.4 WILCOXON RANK SUM TEST WITH THE BONFERRONI ADJUSTMENT

9.6.4.1 The Wilcoxon Rank Sum Test with the Bonferroni Adjustment is a nonparametric test for comparing treatments with a control. The data are ranked and the analysis proceeds exactly as in Steel's Test except that Bonferroni's adjustment for multiple comparisons is used instead of Steel's tables. When Steel's test can be used (i.e., when there are equal numbers of data points per toxicant concentration), it will be more powerful (able to detect smaller differences as statistically significant) than the Wilcoxon Rank Sum Test with Bonferroni's adjustment.

9.6.4.2 The estimate of the safe concentration is reported as the NOEC. A step-by-step example of the use of the Wilcoxon Rank Sum Test with Bonferroni Adjustment is provided in Appendix F.

9.6.5 A CAUTION IN THE USE OF HYPOTHESIS TESTING

9.6.5.1 If in the calculation of an NOEC by hypothesis testing, two tested concentrations cause statistically significant adverse effects, but an intermediate concentration did not cause statistically significant effects, the results should be used with extreme caution.

9.7 **POINT ESTIMATION TECHNIQUES**

9.7.1 PROBIT ANALYSIS

9.7.1.1 Probit Analysis is used to estimate the LC1, LC50, EC1, or EC50 and the associated 95% confidence interval. The analysis consists of adjusting the data for mortality in the control, and then using a maximum likelihood technique to estimate the parameters of the underlying log tolerance distribution, which is assumed to have a particular shape.

9.7.1.2 The assumption upon which the use of Probit Analysis is contingent is a normal distribution of log tolerances. If the normality assumption is not met, and at least two partial mortalities are not obtained, Probit Analysis should not be used. It is important to check the results of Probit Analysis to determine if use of the analysis is appropriate. The chi-square test for heterogeneity provides one good test of appropriateness of the analysis. The computer program (see Appendix I) checks the chi-square statistic calculated for the data set against the tabular value, and provides an error message if the calculated value exceeds the tabular value.

9.7.1.3 A discussion of Probit Analysis, and examples of computer program input and output, are found in Appendix I.

9.7.1.4 In cases where Probit Analysis is not appropriate, the LC50 and associated confidence interval may be estimated by the Spearman-Karber Method (Appendix J) or the Trimmed Spearman-Karber Method (Appendix K). If the test results in 100% survival and 100% mortality in adjacent treatments (all or nothing effect), the LC50 may be estimated using the Graphical Method (Appendix L).

9.7.2 LINEAR INTERPOLATION METHOD

9.7.2.1 The Linear Interpolation Method (see Appendix M) is a procedure to calculate a point estimate of the effluent or other toxicant concentration [Inhibition Concentration, (IC)] that causes a given percent reduction (e.g., 25%, 50%, etc.) in the reproduction or growth of the test organisms. The procedure was designed for general applicability in the analysis of data from short-term chronic toxicity tests.

9.7.2.2 Use of the Linear Interpolation Method is based on the assumptions that the responses (1) are monotonically non-increasing (the mean response for each higher concentration is less than or equal to the mean response for the previous concentration), (2) follow a piecewise linear response function, and (3) are from a random, independent, and representative sample of test data. The assumption for piecewise linear response cannot be tested statistically, and no defined statistical procedure is provided to test the assumption for monotonicity. Where the observed means are not strictly monotonic by examination, they are adjusted by smoothing. In cases where the responses at the low toxicant concentrations are much higher than in the controls, the smoothing process may result in a large upward adjustment in the control mean.

9.7.2.3 The inability to test the monotonicity and piecewise linear response assumptions for this method makes it difficult to assess when the method is, or is not, producing reliable results. Therefore, the method should be used with caution when the results of a toxicity test approach an "all or nothing" response from one concentration to the next in the concentration series, and when it appears that there is a large deviation from monotonicity. See Appendix M for a more detailed discussion of the use of this method and a computer program available for performing calculations.

SECTION 10

REPORT PREPARATION AND TEST REVIEW

10.1 REPORT PREPARATION

The following general format and content are recommended for the report:

10.1.1 INTRODUCTION

1. Permit number
2. Toxicity testing requirements of permit
3. Plant location
4. Name of receiving water body
5. Contract Laboratory (if the tests are performed under contract)
   a Name of firm
   b. Phone number
   c. Address
6. Objective of test

10.1.2 PLANT OPERATIONS

1. Product(s)
2. Raw materials
3. Operating schedule
4. Description of waste treatment
5. Schematic of waste treatment
6. Retention time (if applicable)
7. Volume of waste flow (MGD, CFS, GPM)
8. Design flow of treatment facility at time of sampling

10.1.3 SOURCE OF EFFLUENT, RECEIVING WATER, AND DILUTION WATER

1. Effluent Samples
   a. Sampling point (including latitude and longitude)
   b. Collection dates and times
   c. Sample collection method
   d. Physical and chemical data
   e. Mean daily discharge on sample collection date
   f. Lapsed time from sample collection to delivery
   g. Sample temperature when received at the laboratory

2. Receiving Water Samples
   a. Sampling point (including latitude and longitude)
   b. Collection dates and times
   c. Sample collection method
   d. Physical and chemical data
   e. Streamflow (at time of sampling)
   f. Sample temperature when received at the laboratory
   g Lapsed time from sample collection to delivery

3. Dilution Water Samples
   a. Source

b. Collection date(s) and time(s)
c. Pretreatment
d. Physical and chemical characteristics

10.1.4 TEST METHODS

1. Toxicity test method used (title, number, source)
2. Endpoint(s) of test
3. Deviation(s) from reference method, if any, and the reason(s)
4. Date and time test started
5. Date and time test terminated
6. Type and volume of test chambers
7. Volume of solution used per chamber
8. Number of organisms per test chamber
9. Number of replicate test chambers per treatment
10. Acclimation of test organisms (temperature mean and range)
11. Test temperature (mean and range)
12. Specify if aeration was needed
13. Feeding frequency, and amount and type of food
14. Specify if (and how) pH control measures were implemented

10.1.5 TEST ORGANISMS

1. Scientific name and how determined
2. Age
3. Life stage
4. Mean length and weight (where applicable)
5. Source
6. Diseases and treatment (where applicable)
7. Taxonomic key used for species identification

10.1.6 QUALITY ASSURANCE

1. Reference toxicant used routinely; source
2. Date and time of most recent reference toxicant test, test results, and current control chart
3. Dilution water used in reference toxicant test
4. Results (NOEC or, where applicable, LOEC, LC50, EC50, IC25 and/or IC50); report percent minimum significant difference (PMSD) calculated for sublethal endpoints determined by hypothesis testing in reference toxicant test
5. Physical and chemical methods used

10.1.7 RESULTS

1. Provide raw toxicity data in tabular form, including daily records of affected organisms in each concentration (including controls) and replicate, and in graphical form (plots of toxicity data)
2. Provide table of LC50s, NOECs, IC25, IC50, etc. (as required in the applicable NPDES permit)
3. Indicate statistical methods used to calculate endpoints
4. Provide summary table of physical and chemical data
5. Tabulate QA data
6. Provide percent minimum significant difference (PMSD) calculated for sublethal endpoints

10.1.8 CONCLUSIONS AND RECOMMENDATIONS

1. Relationship between test endpoints and permit limits
2. Actions to be taken

10.2 TEST REVIEW

10.2.1 Test review is an important part of an overall quality assurance program (Section 4) and is necessary for ensuring that all test results are reported accurately. Test review should be conducted on each test by both the testing laboratory and the regulatory authority.

10.2.2 SAMPLING AND HANDLING

10.2.2.1 The collection and handling of samples are reviewed to verify that the sampling and handling procedures given in Section 8 were followed. Chain-of-custody forms are reviewed to verify that samples were tested within allowable sample holding times (Subsection 8.5.4). Any deviations from the procedures given in Section 8 should be documented and described in the data report (Subsection 10.1).

10.2.3 TEST ACCEPTABILITY CRITERIA

10.2.3.1 Test data are reviewed to verify that test acceptability criteria (TAC) requirements for a valid test have been met. Any test not meeting the minimum test acceptability criteria is considered invalid. All invalid tests must be repeated with a newly collected sample.

10.2.4 TEST CONDITIONS

10.2.4.1 Test conditions are reviewed and compared to the specifications listed in the summary of test condition tables provided for each method. Physical and chemical measurements taken during the test (e.g., temperature, pH, and DO) also are reviewed and compared to specified ranges. Any deviations from specifications should be documented and described in the data report (Subsection 10.1).

10.2.4.2 The summary of test condition tables presented for each method identify test conditions as required or recommended. For WET test data submitted under NPDES permits, all required test conditions must be met or the test is considered invalid and must be repeated with a newly collected sample. Deviations from recommended test conditions must be evaluated on a case-by-case basis to determine the validity of test results. Deviations from recommended test conditions may or may not invalidate a test result depending on the degree of the departure and the objective of the test. The reviewer should consider the degree of the deviation and the potential or observed impact of the deviation on the test result before rejecting or accepting a test result as valid. For example, if dissolved oxygen is measured below 4.0 mg/L in one test chamber, the reviewer should consider whether any observed mortality in that test chamber corresponded with the drop in dissolved oxygen.

10.2.4.3 Whereas slight deviations in test conditions may not invalidate an individual test result, test condition deviations that continue to occur frequently in a given laboratory may indicate the need for improved quality control in that laboratory.

10.2.5 STATISTICAL METHODS

10.2.5.1 The statistical methods used for analyzing test data are reviewed to verify that the recommended flowcharts for statistical analysis were followed. Any deviation from the recommended flowcharts for selection of statistical methods should be noted in the data report. Statistical methods other than those recommended in the statistical flowcharts may be appropriate (see Subsection 9.4.1.2), however, the laboratory must document the use of and provide the rationale for the use of any alternate statistical method. In all cases (flowchart recommended

methods or alternate methods), reviewers should verify that the necessary assumptions are met for the statistical method used.

10.2.6 CONCENTRATION-RESPONSE RELATIONSHIPS

10.2.6.1 The concept of a concentration-response, or more classically, a dose-response relationship is "the most fundamental and pervasive one in toxicology" (Casarett and Doull, 1975). This concept assumes that there is a causal relationship between the dose of a toxicant (or concentration for toxicants in solution) and a measured response. A response may be any measurable biochemical or biological parameter that is correlated with exposure to the toxicant. The classical concentration-response relationship is depicted as a sigmoidal shaped curve, however, the particular shape of the concentration-response curve may differ for each coupled toxicant and response pair. In general, more severe responses (such as acute effects) occur at higher concentrations of the toxicant, and less severe responses (such as chronic effects) occur at lower concentrations. A single toxicant also may produce multiple responses, each characterized by a concentration-response relationship. A corollary of the concentration-response concept is that every toxicant should exhibit a concentration-response relationship, given that the appropriate response is measured and given that the concentration range evaluated is appropriate. Use of this concept can be helpful in determining whether an effluent possesses toxicity and in identifying anomalous test results.

10.2.6.2 The concentration-response relationship generated for each multi-concentration test must be reviewed to ensure that calculated test results are interpreted appropriately. USEPA (2000a) provides guidance on evaluating concentration-response relationships to assist in determining the validity of WET test results. All WET test results (from multi-concentration tests) reported under the NPDES program should be reviewed and reported according to USEPA guidance on the evaluation of concentration-response relationships (USEPA, 2000a). This guidance provides review steps for 10 different concentration-response patterns that may be encountered in WET test data. Based on the review, the guidance provides one of three determinations: that calculated effect concentrations are reliable and should be reported, that calculated effect concentrations are anomalous and should be explained, or that the test was inconclusive and the test should be repeated with a newly collected sample. It should be noted that the determination of a valid concentration-response relationship is not always clear cut. Data from some tests may suggest consultation with professional toxicologists and/or regulatory officials. Tests that exhibit unexpected concentration-response relationships also may indicate a need for further investigation and possible retesting.

10.2.7 REFERENCE TOXICANT TESTING

10.2.7.1 Test review of a given effluent or receiving water test should include review of the associated reference toxicant test and current control chart. Reference toxicant testing and control charting is required for documenting the quality of test organisms (Subsection 4.7) and ongoing laboratory performance (Subsection 4.16). The reviewer should verify that a quality control reference toxicant test was conducted according to the specified frequency required by the permitting authority or recommended by the method (e.g., monthly). The test acceptability criteria, test conditions, concentration-response relationship, and test sensitivity of the reference toxicant test are reviewed to verify that the reference toxicant test conducted was a valid test. The results of the reference toxicant test are then plotted on a control chart (see Subsection 4.16) and compared to the current control chart limits (± 2 standard deviations).

10.2.7.2 Reference toxicant tests that fall outside of recommended control chart limits are evaluated to determine the validity of associated effluent and receiving water tests (see Subsection 4.16). An out of control reference toxicant test result does not necessarily invalidate associated test results. The reviewer should consider the degree to which the reference toxicant test result fell outside of control chart limits, the width of the limits, the direction of the deviation (toward increasing test organism sensitivity or toward decreasing test organism sensitivity), the test conditions of both the effluent test and the reference toxicant test, and the objective of the test. More frequent and/or concurrent reference toxicant testing may be advantageous if recent problems (e.g., invalid tests, reference toxicant test results outside of control chart limits, reduced health of organism cultures, or increased within-test variability) have been identified in testing.

10.2.8 TEST VARIABILITY

10.2.8.1 The within-test variability of individual tests should be reviewed. Excessive within-test variability may invalidate a test result and warrant retesting. For evaluating within-test variability, reviewers should consult EPA guidance on upper and lower percent minimum significant difference (PMSD) bounds (USEPA, 2000b).

10.2.8.2 When NPDES permits require sublethal hypothesis testing endpoints from Methods 1000.0,1002.0, or 1003.0 (e.g., growth or reproduction NOECs and LOECs), within-test variability must be reviewed and variability criteria must be applied as described in this section (10.2.8.2). When the methods are used for non-regulatory purposes, the variability criteria herein are recommended but are not required, and their use (or the use of alternative variability criteria) may depend upon the intended uses of the test results and the requirements of any applicable data quality objectives and quality assurance plan.

10.2.8.2.1 To measure test variability, calculate the percent minimum significant difference (PMSD) achieved in the test. The PMSD is the smallest percentage decrease in growth or reproduction from the control that could be determined as statistically significant in the test. The PMSD is calculated as 100 times the minimum significant difference (MSD) divided by the control mean. The equation and examples of MSD calculations are shown in Appendix C. PMSD may be calculated legitimately as a descriptive statistic for within-test variability, even when the hypothesis test is conducted using a non-parametric method. The PMSD bounds were based on a representative set of tests, including tests for which a non-parametric method was required for determining the NOEC or LOEC. The conduct of hypothesis testing to determine test results should follow the statistical flow charts provided for each method. That is, when test data fail to meet assumptions of normality or heterogeneity of variance, a non-parametric method (determined following the statistical flowchart for the method) should be used to calculate test results, but the PMSD may be calculated as described above (using parametric methods) to provide a measure of test variability.

10.2.8.2.2 Compare the PMSD measured in the test with the upper PMSD bound variability criterion listed in Table 6. When the test PMSD exceeds the upper bound, the variability among replicates is unusually large for the test method. Such a test should be considered insufficiently sensitive to detect toxic effects on growth or reproduction of substantial magnitude. A finding of toxicity at a particular concentration may be regarded as trustworthy, but a finding of "no toxicity" or "no statistically significant toxicity" at a particular concentration should not be regarded as a reliable indication that there is no substantial toxic effect on growth or reproduction at that concentration.

10.2.8.2.3 If the PMSD measured for the test is less than or equal to the upper PMSD bound variability criterion in Table 6, then the test's variability measure lies within normal bounds and the effect concentration estimate (e.g., NOEC or LOEC) would normally be accepted unless other test review steps raise serious doubts about its validity.

10.2.8.2.4 If the PMSD measured for the test exceeds the upper PMSD bound variability criterion in Table 6, then one of the following two cases applies (10.2.8.2.4.1, 10.2.8.2.4.2).

10.2.8.2.4.1 If toxicity is found at the permitted receiving water concentration (RWC) based upon the value of the effect concentration estimate (NOEC or LOEC), then the test shall be accepted and the effect concentration estimate may be reported, unless other test review steps raise serious doubts about its validity.

10.2.8.2.4.2 If toxicity is not found at the permitted RWC based upon the value of the effect concentration estimate (NOEC or LOEC) and the PMSD measured for the test exceeds the upper PMSD bound, then the test shall not be accepted, and a new test must be conducted promptly on a newly collected sample.

10.2.8.2.5 To avoid penalizing laboratories that achieve unusually high precision, lower PMSD bounds shall also be applied when a hypothesis test result (e.g., NOEC or LOEC) is reported. Lower PMSD bounds, which are based on the 10^{th} percentiles of national PMSD data, are presented in Table 6. The 10^{th} percentile PMSD represents a practical limit to the sensitivity of the test method because few laboratories are able to achieve such precision on a

regular basis and most do not achieve it even occasionally. In determining hypothesis test results (e.g., NOEC or LOEC), a test concentration shall not be considered toxic (i.e., significantly different from the control) if the relative difference from the control is less than the lower PMSD bounds in Table 6. See USEPA, 2000b for specific examples of implementing lower PMSD bounds.

10.2.8.3 To assist in reviewing within-test variability, EPA recommends maintaining control charts of PMSDs calculated for successive effluent tests (USEPA, 2000b). A control chart of PMSD values characterizes the range of variability observed within a given laboratory, and allows comparison of individual test PMSDs with the laboratory's typical range of variability. Control charts of other variability and test performance measures, such as the MSD, standard deviation or CV of control responses, or average control response, also may be useful for reviewing tests and minimizing variability. The log of PMSD will provide an approximately normal variate useful for control charting.

TABLE 6. VARIABILITY CRITERIA (UPPER AND LOWER PMSD BOUNDS) FOR SUBLETHAL HYPOTHESIS TESTING ENDPOINTS SUBMITTED UNDER NPDES PERMITS.[1]

Test Method	Endpoint	Lower PMSD Bound	Upper PMSD Bound
Method 1000.0, Fathead Minnow Larval Survival and Growth Test	growth	12	30
Method 1002.0, *Ceriodaphnia dubia* Survival and Reproduction Test	reproduction	13	47
Method 1003.0, *Selenastrum capricornutum* Growth Test	growth	9.1	29

[1] Lower and upper PMSD bounds were determined from the 10th and 90th percentile, respectively, of PMSD data from EPA's WET Interlaboratory Variability Study (USEPA, 2001a; USEPA, 2001b).